Recent Results in Extensions to Simultaneous Localization and Mapping

Sanjiv Singh, George Kantor, and Dennis Strelow

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15217, USA

Abstract. We report experimental results with bearings-only and range-only Simultaneous Localization and Mapping (SLAM). In the former case, we give the initial results from a new method that extends optimal shape-from-motion to incorporate angular rate and linear acceleration data. In the latter case, we have formulated a version of the SLAM problem that presumes a moving sensor able to measure only range to landmarks in the environment. Experimental results for both are presented.

1 Introduction

A moving robot can be localized given a map of landmarks and an onboard sensor to sense the landmarks. Conversely, given accurate localization of the sensor, it is possible to build a map of the landmarks. A question that has recently intrigued our community is how well it is possible to do both (localize and map) if neither exist *a priori*. Research in Simultaneous Localization and Mapping (SLAM) makes one of two assumptions. First, and commonly, that the moving sensor can measure both range and bearing to landmarks in its vicinity[4], or, that the moving sensor can detect only bearing to or projections of the landmarks[1][9]. Interestingly enough, little attention has been paid to the complementary case, in which the moving sensor can only detect range to landmarks.

For the case in which projections are sensed, we have extended optimal methods for shape-from-motion to incorporate angular rate and linear acceleration data from rate gyros and accelerometers, respectively. In addition to the obvious advantage of redundant measurements, the complimentary nature of visual and inertial data provides the following qualitative advantages over using visual or inertial data alone:

- Visual information can correct the drift that results from integrating inertial sensor data
- Inertial information can disambiguate the camera's motion when visual data is scarce or when the scene viewed by the camera is degenerate
- Visual information can be used to determine the gravity direction required to correctly extract accelerations from accelerometer readings

- Conversely, because the accelerometer readings are affected by gravity, two absolute components of the sensor's orientation can be determined if the contribution of gravity to the reading can be separated from the contribution of acceleration

Here we present an algorithm for estimating camera motion using both visual and inertial data.

For range-only SLAM we have adapted the well-known estimation techniques of Kalman filtering, Markov methods, and Monte Carlo localization to solve the problem of robot localization from range-only measurements[3]. All three of these methods estimate robot position as a distribution of probabilities over the space of possible robot positions. In the same work we presented an algorithm capable of solving SLAM in cases where approximate *a priori* estimates of robot and landmark locations exist. However, a solution to the range-only SLAM problem with no prior information remains to be found. The primary difficulty stems from the annular distribution of potential relative locations that results from a range only measurement Since the distribution is highly non-Gaussian, SLAM solutions based on Kalman filtering falter. In theory, Markov methods (probability grids) and Monte Carlo methods (particle filtering) have the flexibility to handle annular distributions. Unfortunately, the scaling properties of these methods severely limit the number of landmarks that can be mapped.

In truth, Markov and Monte Carlo methods have much more flexibility than we need; they can represent arbitrary distributions while we need only deal with well structured annular distributions. What is needed is a compact way to represent annular distributions together with a computationally efficient way of combining annular distributions with each other and with Gaussian distributions. In most cases, we expect the results of these combinations to be well approximated by mixtures of Gaussians so that standard techniques such as Kalman filtering or multiple hypothesis tracking could be applied to solve the remaining estimation problem.

Here we present new results in our ongoing efforts towards the solution to the general range-only SLAM problem. The key is a computationally efficient method of representing annular distributions and approximating their multiplication. This makes it possible for the robot to estimate locations of new landmarks that are completely unknown to start.

2 Method

2.1 Optimal motion estimation using image and inertial data

In this section we present a method for estimating camera motion and sparse scene structure using image, rate gyro, and accelerometer measurements. The method is a batch algorithm that finds optimal estimates by minimizing a total error with respect to all of the unknown parameters simultaneously.

Error function. The error we minimize is $E = E_{\text{visual}} + E_{\text{inertial}}$, where

$$E_{\text{visual}} = \sum_{i,j} D(\pi(C_{\rho_i,t_i}(X_j)) - x_{ij}) \tag{1}$$

and

$$\begin{aligned} E_{\text{inertial}} = & \sum_{i=1}^{f-1} D\left(\rho_i - I_\rho(\tau_{i-1}, \tau_i, \rho_{i-1})\right) \\ & + \sum_{i=1}^{f-1} D\left(v_i - I_v(\tau_{i-1}, \tau_i, \rho_{i-1}, v_{i-1}, g)\right) \\ & + \sum_{i=1}^{f-1} D\left(t_i - I_t(\tau_{i-1}, \tau_i, \rho_{i-1}, v_{i-1}, g, t_{i-1})\right) \end{aligned} \tag{2}$$

E_{visual} specifies an image reprojection error given the six degree of freedom camera positions and three-dimensional point positions. This error function is similar to those used in bundle adjustment[11] and nonlinear shape-from-motion[10]. In this error, the sum is over i and j, such that point j was observed in image i. x_{ij} is the observed projection of point j in image i. ρ_i and t_i are the camera-to-world rotation Euler angles and camera-to-world translation, respectively, at the time of image i, and C_{ρ_i,t_i} is the world-to-camera transformation specified by ρ_i and t_i. X_j is the world coordinate system location of point j, so that $C_{\rho_i,t_i}(X_j)$ is location of point j in camera coordinate system i. π gives the image projection of a three-dimensional point specified in the camera coordinate system. In our current implementation, π may be either a conventional (i.e., perspective or orthographic) or an omnidirectional projection. The details of our omnidirectional projection model and its use with nonlinear shape-from-motion are given in [9][8].

E_{inertial} gives an error between the estimated positions and velocities and incremental positions and velocities predicted by the inertial measurements. Here, f is the number of images, and τ_i is the time image i was captured. ρ_i and t_i are the camera rotation and translation, just as in the equation for E_{inertial} above. v_i gives the camera's linear velocity at time τ_i, and g is the direction of gravity relative to the camera rotation at time τ_0. I_ρ, I_v, and I_t integrate the inertial observations to produce estimates of ρ_i, v_i, and t_i from initial values ρ_{i-1}, v_{i-1}, and t_{i-1}, respectively.

We assume that all of the individual error functions D are Mahalanobis distances. The image error covariances may be assumed to be uniform and isotropic (e.g., with a standard deviation of one pixel) or determined by the tracking algorithm using image texture[2]. In the experiments below, we have used isotropic densities as tuning parameters to specify relative confidences in the image, rate gyro, and accelerometer observations.

Estimation. We use Levenberg-Marquardt[6] to minimize the combined error with respect to the camera rotations ρ_i, translations t_i, velocities v_i, the

gravity direction g, and the world point locations X_j. The inverse of the Hessian matrix from the Levenberg-Marquardt method provides an estimate of the covariance of the recovered parameters.

2.2 A Geometric Method for Combining Non-Gaussian Distibutions

In [3] we presented a Kalman filter based algorithm capable of solving the range–only SLAM problem in the case where the approximate locations of the landmarks are known *a priori*. In this approach, the system state vector was defined to contain robot pose together with the positions of all of the landmarks. At each time step, range measurements were "linearized" about the current state estimate to produce and approximation of the relative positions between the robot and each of the landmarks. These approximations were then fed into a Kalman filter to improve the state estimate. Here we extend these results to allow a robot to initialize new landmarks that are completely unknown. The basic idea is to store the robot locations and measured ranges the first few times a landmark is encountered and then estimate its position by intersecting circles on the plane. Once an estimate of a new landmark is produced, it is added to the Kalman filter where its estimate is then improved along with the estimates of the other (previously seen) landmarks. The keys to this idea are to (1) find the intersection points of two circles and (2) estimate a distribution about the intersection points. Because it takes advantage of the special structure of the problem, the resulting approach is less computationally cumbersome and avoids the local extrema problems associated with standard batch optimization techniques.

Merging Annular Distributions. Given two circles with centers $p_1 = [x1, y1]^T$ and $p_2 = [x2, y2]^T$, and radii r_1 and r_2, respectively, [1] their two points of intersection are given by:

$$p = p_m \pm \begin{bmatrix} (y_2 - y_1)\frac{\sqrt{r_1^2 - a^2}}{d} \\ -(y_2 - y_1)\frac{\sqrt{r_1^2 - a^2}}{d} \end{bmatrix}, \tag{3}$$

where where $d = \|p_2 - p_1\|$,

$$a = \frac{r_1^2 - r_2^2 + d^2}{2d}, \qquad \text{and} \qquad p_m = p_1 + a\frac{p_2 - p_1}{d}.$$

Next we approximate the distribution around each point with a Gaussian. At each intersection point, we obtain two Gaussian approximations, one for each annulus, following the procedure outlined in [3]. We then merge the two approximations into a single Gaussian using standard Kalman gain formulas

[1] In our application, p_1 and p_2 are the robot locations at two different times, and r_1 and r_2 are the associated measurements.

(see eg. Smith and Cheeseman[7]). This process is depicted graphically in Figure 1. Since there will usually be two intersection points, this process usually results in two Gaussian distributions, one for each intersection point. In the absence of any other information, we weight each Gaussian equally. The result is that this pair of Gaussians forms an approximation of the distribution that results from multiplying two annular distributions.

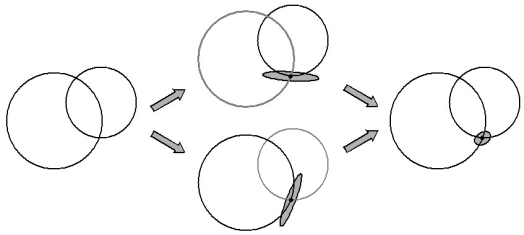

Fig. 1. Approximating intersections of annular distribution as Gaussian. First, a Gaussian approximation is determined for each annulus about the intersection point. Then the two individual approximations are merged into one.

To approximate the distribution that results from multiplying three annular distributions, we iterate pairwise intersection in the following manner. An example of this process is depicted in Figure 2. Let (p_1, C_1) and (p_2, C_2) be the means and covariance matrices associated with the two Gaussians used to approximate the multiplication of the first two annuli. Likewise, let (q_1, D_1) and (q_2, D_2) denote the means and covariances associated with multiplication of the first and third annuli. We then create four new distributions, (p_{11}, C_{11}), (p_{12}, C_{12}), (p_{21}, C_{21}), and (p_{22}, C_{22}), where (p_{ij}, C_{ij}) are the mean and covariance matrix that result from merging (p_i, C_i) and (q_j, D_j). We assign a weight w_{ij} to each of these four distributions that is inversely proportional to the distance between p_i and q_j. We then eliminate distributions whose weights are below some threshold and rescale the remaining weights. Now we rename the remaining distributions to be (p_1, C_1) through (p_{n_1}, C_{n_1}), where n_1 is the number of remaining distributions. We then introduce the Gaussians that result from merging the second and third annuli, and we name their means and covariances (q_1, D_1) and (q_2, D_2). We then follow a similar process of creating $2n_1$ new distributions by merging each (p_i, C_i) with each (q_j, D_j), weighting each new distribution inversely with the distance between p_i and q_j, throwing out distributions whose weights are below a certain threshold, and renaming the remaining distributions (p_1, C_1) through (p_{n_2}, C_{n_2}), where n_2 is the number of remaining distributions. The resulting set of distributions together with their weights give an approximation of the distribution that results from multiplying three annular distributions. Often n_2 is equal to one, and the approximation is a single Gaussian.

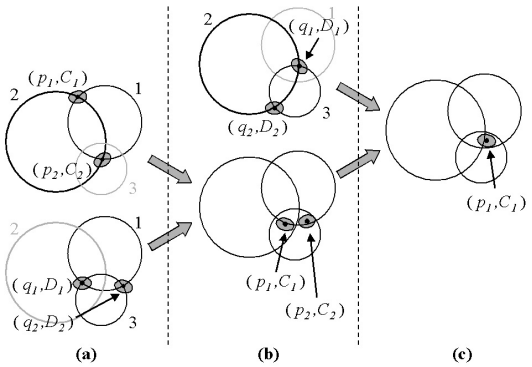

Fig. 2. Gaussian approximation for three intersecting annuli. In (a) Gaussian approximations resulting from the intersection of the first and second annuli (top) and the first and third annuli (bottom) are combined to form the distributions in (b, bottom). All possible combinations of individual Gaussians are considered and unlikely combinations are thrown away. In this example two Gaussians remain after the first step, but there could be as many as four or as few as one. In (b) the Gaussian approximations resulting from the intersection of second and third annuli are combined with the result of the first step. The final result, shown in (c), is a single Gaussian for this example.

The procedure for approximating the multiplication of four or more annuli begins by finding the approximation for the first three then follows a similar pattern of finding pairwise approximations, merging, weighting, and truncating.

Estimating Landmark Location. Recall that our approach to estimating the location of a newly acquired landmark is to store the robot locations and range measurements the first few times the landmark is encountered. With this in mind, let n be the number of measurements to the new landmark and for $i \in \{1, 2, \ldots, n\}$ let $((x_i, y_i), P_i)$ be the robot location estimate (mean and covariance matrix) and let r_i be the range measurement the ith time that the new landmark is encountered. The best estimate for the location of the new landmark is given by the location of the peak of the distribution that results from multiplying the n annuli with centers (x_i, y_i) and radii r_i, $i = 1, 2, \ldots, n$. To get an approximation of this estimate, we use the algorithm described above to approximate the distribution that results from multiplying the n annuli with a weighted collection of Gaussians and choose the estimate to be the mean of the Gaussian with the highest weight.

This approach requires estimates of the robot positions corresponding to the measurements to the new landmarks. In the experimental results presented in Section 3.2, we obtain these estimates by using odometry measurements together with measurements to other landmarks that have already been

mapped. When no landmark locations are known in advance, this method can still be used to initialize new landmarks using only odometry data.

3 Results

3.1 Optimal motion estimation using image and inertial data

In this section we present an experimental result from our algorithm for optimal motion estimation using both image and inertial measurements. In this experiment, we have mounted a conventional perspective camera augmented with rate gyros and an accelerometer on a five degree of freedom manipulator, which provides repeatable and known motions. More details on the configuration and observations are given below, along with an analysis of the resulting motion estimates.

Configuration. Our sensor rig consists of an industrial vision camera paired with a 6 mm lens, three orthogonally mounted rate gyros, and a three degree of freedom accelerometer. The camera exposure time is set to 1/200 second. The gyros and accelerometer measure motions of up to 150 degrees per second and 4 g, respectively.Images were captured at 30 Hz using a conventional frame grabber. To remove the effects of interlacing, only one field was used from each image, producing 640×240 pixel images. Voltages from the gyros and the accelerometer were simultaneously captured at 200 Hz.

The camera intrinsic parameters (e.g., focal length and radial distortion) were calibrated using the method in [5]. This calibration also accounts for the reduced geometry of our one-field images. The accelerometer voltage-to-acceleration calibration was performed using a field calibration that accounts for non-orthogonality between the individual accelerometers. The individual gyro voltage-to-rate calibrations were determined using a turntable with a known rotational rate. The fixed gyro-to-camera and accelerometer-to-camera rotations were assumed known from the mechanical specifications of the mount.

Observations. To perform experiments with known and repeatable motions, we mounted our rig on a robotic arm. In our experiment, the camera points toward the ceiling, and translates in x, y, and z through seven pre-specified points, for a total distance traveled of about 2.0 meters. Projected onto the (x, y) plane, these points are located on a square, and the camera moves on a curved path between points, producing a clover-like pattern in (x, y). The camera rotates through an angle of 270 degrees about the camera's optical axis during the course of the motion.

We tracked 23 features though a sequence of 152 images using the Lucas-Kanade algorithm. A few images from the sequence are shown in Figure 3. As shown in the figure, only 5 or 6 of the 23 features are typically visible in any one image. Because the sequence contains repetitive texture and large interframe motions, mistracking was common and was corrected manually.

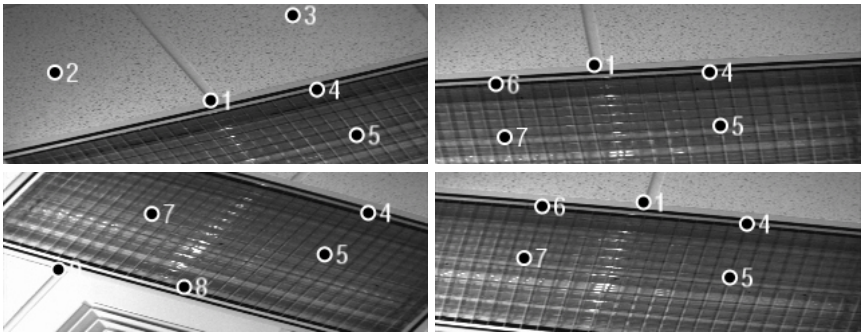

Fig. 3. Images 16, 26, 36, and 46 from the test sequence are shown clockwise from the upper left. As described in section 3.1, the images are one field of an interlaced image, so their height is half that of the full image.

Estimated motion. Some aspects of the estimated structure and motion are shown graphically in Figures 4-6. Figure 4 shows the (x, y) translation estimated using image measurements only (the erratic solid line) and both image and inertial measurements (the dash-dotted line). This figure also shows the motion estimate that results from integrating the inertial data only, assuming zero initial velocity and using the optimal gravity estimate (the diverging dashed line). Random errors in the measured accelerations and a small error in the estimated gravity cause this path to diverge almost immediately from the correct path, and accumulated error in the integrated velocity soon causes gross errors in the motion estimate.

Figure 5 shows the estimated error covariances for the (x, y) translations for every fifteenth image of the sequence. The covariances resulting from using image measurements only are shown as the large dotted ellipses, which show the one standard deviation error boundaries. The error covariances resulting from using both image and inertial measurements are shown as the solid ellipses, which show five standard deviation error boundaries. To provide a direct comparison, the covariances for both cases are estimated at the solution found using both image and inertial measurements. This solution, shown as a dash-dotted line, is the same as in Figure 2.

Similarly, Figure 6 shows the estimated error covariances for the (x, y) point locations. As in Figure 5, the estimated error covariances were evaluated at the solution found using both image and inertial measurements, which is shown as a dash-dotted line for reference.

3.2 Initializing Unknown Landmarks

In this section we present the results of a range-only tracking experiment during which a new, completely unknown landmark is initialized and estimated.

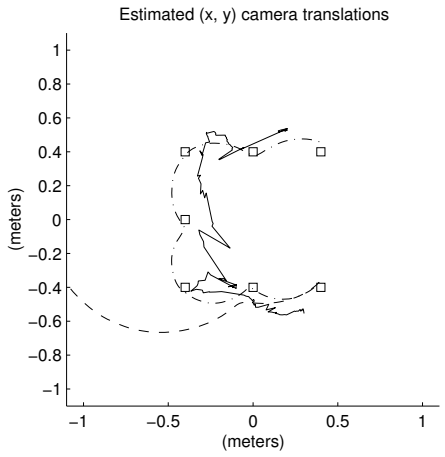

Fig. 4. The (x, y) camera translations estimated using only image measurements, only inertial measurements, and both image and inertial measurements are shown as the erratic solid path, smooth diverging dashed path, and dash-dotted path, respectively. Seven known points on the camera's actual path are shown as squares.

Experimental Configuration. This experiment employed an electric all-terrain vehicle (ATV) equipped with an inertial navigation system (INS) and a range finding radio-frequency identification (RFID) system. The INS system consists of a fiber optic gyro to provide the vehicle's angular velocity together with an odometer to measure distance traveled. Outputs from the gyro and odometer are integrated to give a dead-reckoning estimate of position.

In the system used, a cell controller attached to an antenna continually sends queries out into the world. When the query is received by a credit-card-sized radio transponder (tag), the tag responds with its unique identification number. When the tag response is received back at the antenna, the cell controller uses round trip time of flight to estimate the range between the antenna and tag. In our experiment, a cell controller and a collection of antennas configured to provide a 360 degree coverage pattern are mounted onboard the ATV. Ten tags, which serve as landmarks for range-only tracking and SLAM, are distributed over a planar, unobstructed environment. The layout of the tags is depicted in Figure 8.

Tracking Results. Figure 7 shows the results of a tracking experiment where range-only measurements are used to correct for errors in the INS dead-reckoning position estimate. At the beginning of the experiment, the positions of nine tags are known exactly while the position of the tenth tag is completely unknown. The figure plots actual robot position (dotted line), INS dead-reckoning position estimate (dashed line), and a radio-tag tracking estimate based on extended Kalman filtering (solid line) for a portion of

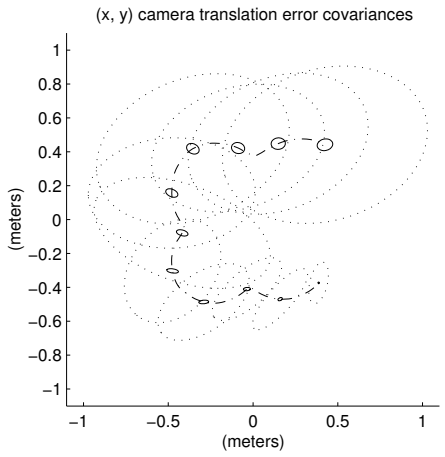

Fig. 5. The estimated error covariances of the (x, y) camera translation for every fifteenth image of the sequence. The one standard deviation boundaries that result from using only image measurements and the five standard deviation boundaries that result from using both image and inertial measurements are shown as the dotted and solid ellipses, respectively.

the experiment. Here, the initial position of the estimate was chosen to be the same as the actual initial position, with an initial orientation error of 45 degrees. Because it does not use any outside information, the INS estimate cannot recover from this initial error.

Range data is received somewhat infrequently. During the 225 second experiment, a total of 107 range measurements were received from the nine tags. This explains the erratic nature of the improved estimate; there are long periods of time during which no range measurements are received and hence the estimate error grows. When a measurement is received, the resulting correction can appear drastic

Mapping Results. During the tracking experiment described above, the position of a tenth, completely unknown tag was initially estimated using the algorithm described in Section 2.2. Once an initial estimate was produced, it was improved using the range-only SLAM algorithm presented in [3]. The results are shown in Figure 7. The plot shows both the initial estimate generated using circle intersection (solid ellipse) and the final estimate improved by application of Kalman filter based SLAM (dotted ellipse).

References

1. M. Deans and M. Hebert. Experimental comparison of techniques for localization and mapping using a bearing–only sensor. In Daniela Rus and Sanjiv Singh, editors, *Experimental Robotics VII*, pages 393–404. Springer-Verlag.

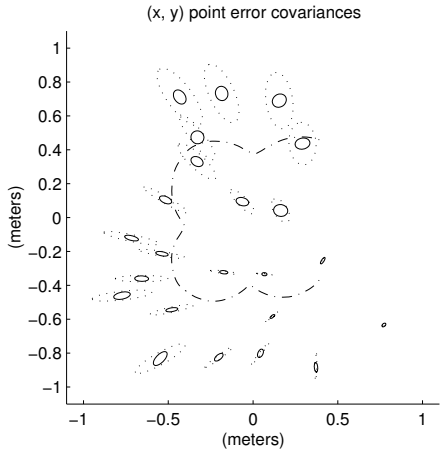

Fig. 6. The estimated error covariances of the (x, y) point positions. As in Figure 5, the one standard deviation boundaries that result from using only the image measurements and the the five standard deviation boundaries that result from using both image and inertial measurements are shown as dotted and solid ellipses, respectively.

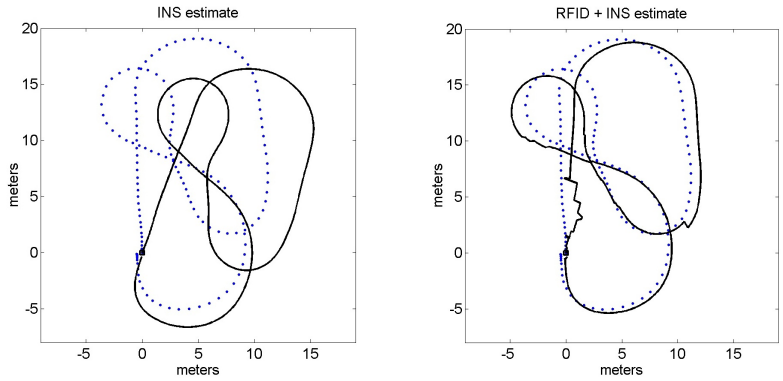

Fig. 7. Actual and estimated robot trajectories for a portion of the range-only tracking and SLAM experiment. In both figures, the actual robot trajectory (ground truth) is plotted with a dotted line. The figure on the left shows the estimate generated using dead-reckoning and INS sensors. The Kalman filter tracking estimate that fuses INS output and radio tag range measurements is shown in the figure on the right.

2. Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Proceedings of the Eighth International Conference on Computer Vision*, Vancouver, Canada, 2001.

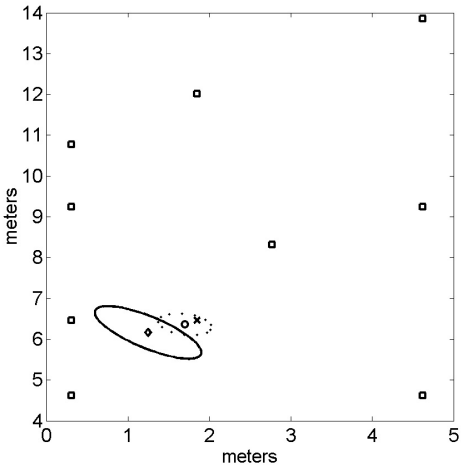

Fig. 8. Initial and final estimates of unknown landmark. The diamond and associated ellipse (solid line) represent the estimate and covariance resulting from the batch initialization algorithm described in Section 2.2. The circle and associated ellipse (dotted line) represent the final estimate, improved by Kalman filter SLAM algorithm, at the end of the experiment. The actual position of the unknown landmark is marked with an x, and the squares denote the positions of the previously mapped landmarks.

3. G. Kantor and S. Singh. Preliminary results in range–only localization and mapping. In *Proceedings of ICRA 2002*, pages 1819–1825, May 2002.
4. J.J. Leonard and H. Feder. A computationally efficient method for large-scale concurrent mapping and localization. In *Robotics Research: The Ninth International Symposium*, pages 169–176, Snowbird,UT, 2000. Springer Verlag.
5. Intel Corporation. Open source computer vision library. `http://www.intel.com/research/mrl/ research/cvlib/`.
6. William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992.
7. R.C. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research*, 5(4):56–68, 1986.
8. Dennis Strelow, Jeff Mishler, David Koes, and Sanjiv Singh. Precise omnidirectional camera calibration. In *IEEE Computer Vision and Pattern Recognition*, Kauai, Hawaii, December 2001.
9. Dennis Strelow, Jeff Mishler, Sanjiv Singh, and Herman Herman. Extending shape-from-motion to noncentral omnidirectional cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, Hawaii, October 2001.
10. Richard Szeliski and Sing Bing Kang. Recovering 3D shape and motion from image streams using nonlinear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, March 1994.
11. Paul R. Wolf. *Elements of Photogrammetry*. McGraw-Hill, New York, 1983.